Experimental genomics: The application of DNA microarrays in cellular and molecular biology studies

LUO Xiao-yan [1], TANG Wei [2]

([1]*Department of Cell and Developmental Biology, University of North Carolina, Chapel Hill, NC 27599, USA*)

([2]*Forest Biotechnology Group, North Carolina State University, Centennial Campus, P.O.Box 7247, Raleigh, NC 27695, USA*)

Abstract The genome sequence information in combination with DNA microarrays promises to revolutionize the way of cellular and molecular biological research by allowing complex mixtures of RNA and DNA to interrogated in a parallel and quantitative fashion. DNA microarrays can be used to measure levels of gene expression for tens of thousands of gene simultaneously and take advantage of all available sequence information for experimental design and data interpretation in pursuit of biological understanding. Recent progress in experimental genomics allows DNA microarrays not simply to provide a catalogue of all the genes and information about their function, but to understand how the components work together to comprise functioning cells and organisms. This brief review gives a survey of DNA microarrays technology and its applications in genome and gene function analysis, gene expression studies, biological signal and defense system, cell cycle regulation, mechanism of transcriptional regulation, proteomics, and the functionality of food component.

Keywords: Experimental genomics; Sequence information; DNA microarrays; Gene expression; Functional analysis.

CLC number: Q789 **Document code:** A **Article ID:** 1007-662X(2002)04-0299-10

Introduction

The DNA microarray technology is a new and powerful technology that will substantially increase the speed of cellular and molecular biological research. The DNA microarray technology was invented in Stanford (Schena *et al.*1995) using a small set of Arabidopsis (*Arabidopsis thaliana*) ESTs and has been applied to several model organisms including colon bacillus (*Escherichia coli*) (Richmond *et al.* 1999), yeast (*Saccharomyces cerevisiae*) (Chu *et al.*1998), fruit fly (*Drosophila melanogaster*)(White *et al.* 1999), nematode (*Caenorhabditis elegans*) (The *C. elegans* Sequencing Consortium 1998), mouse (Tanaka *et al.* 2000) and human (Schena *et al.*1996). The major advance on DNA microarray technology, as compared to conventional techniques, results from the small size of the array, which allows for a higher sensitivity, enables the parallel screening of larger numbers of genes and provides the opportunity to use smaller amounts of starting material. The introduction of fluorescent probes has made miniaturization of arrays possible (Jordan *et al.* 1998). The scale of gene expression analysis is not only extended by the simultaneous analysis of large numbers of genes, but also because microarrays can be produced in series facilitating comparative analysis of a large number of samples. Analysis of gene expression is important in many fields of biological research, since changes in the physiology of an organism or a cell will be accompanied by changes in the pattern of gene expression. Gene expression analysis can be used to obtain insight in the physiological consequences of genetic modification in animals and plants. Several techniques for the analysis of gene expression at the mRNA-level, such as Northern blotting (Alwine *et al.* 1997), dot blot analysis (Lennon *et al.* 1991), differential display (Liang and Pardee 19992), and serial analysis of gene expression (SAGE) (Velculescu *et al.* 1995) are available. These methods have their disadvantages, which render them unsuitable if large numbers of expression products have to be analyzed simultaneously. Northern blot analysis only allows limited numbers of mRNAs to be studied at the same time. Dot blot analysis requires a relatively large amount of material due to the size of the filters. Differential display does enable the simultaneous detection of multiple differences in gene expression and screening is based on differences in mRNA length and not identity. SAGE involves complex sample preparation procedures, requires extensive DNA sequencing and is not very sensitive. Recently, substantial improvement in sensitivity and throughput of expression screening has been obtained by the introduction of DNA microarray technology (Watson *et al.* 1998; Duggan *et al.* 1999; Graves 1999).

The study of gene expression by DNA microarray technology is based on hybridization of mRNA to a high-density array of immobilized target sequences, each corresponding to a specific gene. Sample mRNAs are labeled as a complex mixture by incorporation of a fluorescent nucleotide by oligo (dT)-primed reverse transcription. The labeled pool of sample mRNAs is subsequently hybridized to the array, where each messenger will quantitatively hybridize to its complementary target sequence. The fluorescence at

Biography: LUO Xiao-yan (1966-): Postdoctoral research associate in cell and developmental biology and DNA microarray in Univ. of North Carolina at Chapel Hill. E-mail: wtang@unity.ncsu.edu

Received date 2002-05-27

Responsible editor: Chai Ruihai

each spot on the array is a quantitative measure corresponding to the expression level of the particular gene. The use of two differently labeled mRNA samples allows quantitative comparison of gene expression in both samples (Fig. 1) (van Hal *et al.* 2000).

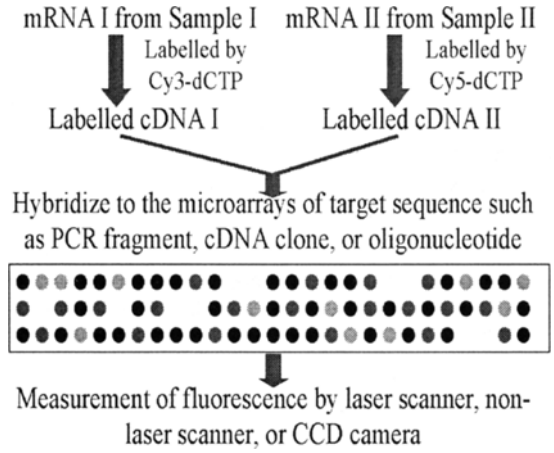

Fig. 1. The principle of gene expression analysis by DNA microarray technology

mRNA samples are reverse transcribed to cDNA while fluorescently labeled nucleotides are incoporated (Cy3 and Cy5 labelled dCTP or dUTP are often used for this purpose). The usage of multiple dyes allows the comparison of multiple RNA samples on one single array.

Currently, two complementary techniques are available. One is fragment-based DNA microarrays, and another is oligonucleotide-based Affymetrix chips. DNA microarrays allow the simultaneous hybridization of two fluorescently labeled probes to an array of immobilized DNA fragments such as PCR-amplified DNA sequences, each corresponding to a specific gene. After scanning of the microarray with a laser scanner, the signal for each DNA fragment reflects the abundance of the corresponding messenger RNA in the sample. The use of two differently labeled samples allows the quantitative comparison of gene expression between a control and a test experiment. Alternatively, Affymetrix chips consist of an array of oligonucleotides of usually 20–25 bp, which have been synthesized in situ on a glass surface using photolithography. Each gene to be analyzed is typically represented by twenty specific probes on the chip (Fig. 1). Different methods for labeling RNA are available and allow a quantitative measurement of transcript abundance. As opposed to fragment-based microarrays, oligonucleotide arrays require prior knowledge of DNA sequence information but permit single base change analysis. DNA microarray technology could be useful in: (1) genome and gene function analysis; (2) risk assessment in transgenic agricultural products by analysis of altered gene expression; (3) studies on biological signal and defense system; (4) un-raveling gene function and metabolic pathways in animals and plants by mutant analysis; (5) cell cycle and transcriptional regulation research; (6) protein-protein interaction; (7) investigation of functional and toxicological effects of food components. Although DNA microarray technology can be widely used for detection and identification of complex samples in most areas of current biology, this paper gives a survey of DNA microarray technology and its use in the important subjects of cellular and molecular biology involved in experimental genomics.

Manufacturing of DNA microarrays

At present, two main techniques are being developed for manufacturing DNA microarrays (Fig. 2). The first approach, DNA chip, encompasses direct synthesis of oligonucleotides on a solid surface based on photolithographyas that was developed by Fodor *et al.* (1991). Specific areas of a glass surface that derivatised with linker molecules that carry a photo-labile protective group, are selectively illuminated by using a photo-mask. Subsequently, the surface is incubated with a solution containing a photo-protected nucleotide, which will only be coupled to the light-activated areas. After removal of the excess nucleotide, a second photo-mask is used to de-protect other areas on the surface and subsequently another type of nucleotide is coupled to these areas. By repeating this procedure, a defined set of oligonucleotides is synthesized on the surface (Fig. 2) (Chee *et al.* 1996; McGall *et al.* 1996; Lipshutz *et al.* 1995). The method allows the manufacturing of microarrays with very high densities at 250, 000 oligonucleotide spots per cm^2 and facilitates the production of large series of identical arrays. However, it is prohibitively expensive and has no flexibility in design. The second approach, DNA micro-dispenser, is more flexible and can be performed in a regular molecular biology laboratory (Schena *et al.* 1995). Small quantities of DNA solution, with a minimum volume of approximately 50pl, are dispensed onto a solid surface. The number of micro-dispensing robots commercially available is quickly increasing and the performance of these machines is continually improving. DNA micro-dispensers apply the DNA solution with a pin that touches the solid surface (Fig. 2). The density of the spots depends on the skills of the dispensing device. DNA dispensing is flexible and allows for constant update of the array. Oligonucleotides as well as longer known or unknown DNA sequences can be deposited on the array in a format that can vary, if necessary, from one array to the next. Using micro-dispenser it should also be possible to synthesize oligonucleotides directly on an array or to deposit molecules other than nucleic acids, for example proteins.

DNA molecules that carry a 5% modification can be covalently bound to a glass surface that carries reactive groups (Schena *et al.* 1996; Rogers *et al.* 1999). Presently surface modified glass slides (silylated, poly-L-lysine coated) are most commonly used as a substrate. Besides glass, other materials are being explored as well, such as gold-coated slides, polyacrylamide gel pads, and nitrocellu-

lose or nylon membranes (Proudnikov *et al.* 1998). These alternatives aim at better signal to noise ratios and improved reproducibility. DNA microarray technology is rapidly evolving in many aspects. To analyze gene expression in animal and plant biopsies, micro-dissected specimens, or small plant tissues, the overall sensitivity has to be increased. This will depend on technical aspects such as the quality of the scanners and arrayers, fluorescent dyes with improved quantum yields and lower background, or target supports with reduced background and more target sequence binding capacity (Wittung *et al.* 1996; Geiger *et al.* 1998). The elucidation of metabolic pathways or identification of novel responding genes necessitates the use of large arrays often with undefined target sequences. Compared to conventional methods, DNA microarrays have the advantage that expression of large sets of genes can be determined in parallel. (Spellman *et al.* 1998; Iyer *et al.* 1999). If the number of genes of an organism is not too large and all the genes are known, sequences corresponding to all open reading frames can be spotted on the array that allows simultaneous experimental analysis of all mRNAs.

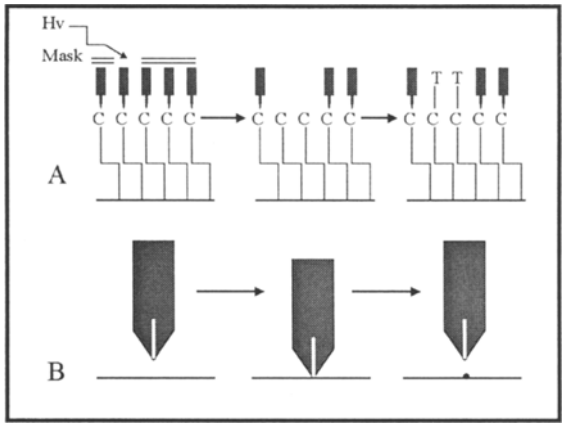

Fig. 2 Two approaches for DNA microarray manufacturing.
A-- Direct synthesis of oligonucleotides on a solid surface based on photolithographyas; B-- DNA micro-dispensers (van Hal *et al.* 2000).

Hybridization, scanning, and data analysis

After mRNA was purified from the biological sample, mRNA should be labeled fluorescently by incorporation of a modified nucleotide during cDNA synthesis prior to hybridization (Fig. 1). Criteria for selecting a fluorophore are a narrow excitation and emission peak, a high level of photon-emission, resulting in better sensitivity, and resistance to photo-bleaching. Presently, the fluorescent cyanine dyes Cy3 and Cy5 are most often used. Cy5 has the disadvantage that it sometimes gives high background fluorescence on glass surfaces and is more sensitive to photo-bleaching than Cy3. The development of fluorescent dyes with improved characteristics and compatible scanners will certainly facilitate a further increase in the signal to noise ratio. The ability to use multiple dyes in a single experiment will allow better comparison of several different mRNA samples simultaneously. The required hybridization conditions such as sample concentration, ionic strength, temperature largely depend on the size of the DNA fragments present on the array and must be determined for a given experimental set up. Although existing sequence databases are of great value for selecting relevant genes, the identification of unknown genes dictates the use of cDNA libraries as source. The correlation between transcript concentration and hybridization signal can be determined by spiking control RNA transcripts in a background of a total messenger population. Work of Lemieux *et al.* (1998) has indicated that the RNA expression level and hybridization signal show a linear correlation providing that the target DNA is present on the array in at least a ten-fold excess.

For imaging, there are several types of microarray readers are available. They can be subdivided into CCD cameras, non-confocal laser scanner sand confocal laser scanners. CCD cameras allow fast scanning but their imaging area is rather small. Confocal laser scanners have the advantage that the light collection efficiency and resolution is usually much higher than for the other two systems. In addition, confocal laser scanners have a small depth of focus that reduces artifacts but also requires more scanning precision. In general, a hybridized microarray was scanned with a confocal laser scanner ScanArray3000 (General Scanning) at 543 nm (Cy3, GHeNe laser) or 632 nm (Cy5, RHeNe laser) repeatedly. Laser power was at 75%. To obtain a reliable quantitative result the microarray spot diameter should be at least five to ten times the pixel size, which is at this moment about 10 mm for confocal laser scanners. Hybridization may be improved by developing new approaches, such as the use of electric fields to enhance the hybridization rate and precisely regulate the stringency of the hybridization (Edman *et al.* 1997; Sosnowski *et al.* 1997).

The complex hybridization data are generated within a short time and sophisticated software is needed to keep a good overview, to assess the quality of the data and to help find statistical significance and relevant correlations within, and between different arrays and experiments. Linking microarray-derived gene expression data to DNA sequence information and experimental data is available in public databases. For example, one can find genes that are part of the same metabolic route, compare their expression behavior under all the tested conditions and present the data in an orderly and visual way. The software should be able to trace back a spot on an array to a clone in the freezer and its sequence in a database. These data handling software packages are commercially available but large improvements are still necessary, especially with regard to spot location, expression pattern recognition, and comparison of larger numbers of different experimental samples, statistical analysis and visual data presentation.

Applications of DNA microarrays

Genome and gene function analysis

Functional analysis, through parallel expression monitoring, should help researchers better understand the fundamental mechanisms that underlie plant growth and development. By accumulating databases of expression information as a function of tissue type, developmental stage, hormone and herbicide treatment, genetic background and environmental condition, it should be possible to identify the genes involved many aspects of current biology. Microarray analysis provides a way to link genomic sequence information and functional analysis. Recent experiments involving the use of cDNA microarrays for expression monitoring indicated the immediate applicability of DNA chips in agricultural biotechnology. The DNA microarray technology may be used to collect much of the data that are obtained presently by Southern and northern hybridization approaches in a more highly parallel fashion. Genomic DNA samples can be manipulated experimentally to select for particular regions before hybridization to obtain specific types of information. The applications of DNA microarrays to genomic studies primarily involve the search for single-nucleotide polymorphisms, which may have considerable importance (Cheung *et al.* 1999). According to defect and the best treatment option, one can use analytical techniques such as genetic-linkage mapping or association analysis (Baldwin *et al.* 1999) to discover genetic predispositions to disease, and to classify diseases. When the locations of nearly all of the specific defects have been determined, the work will become quite a bit easier, because a specific subset of probes can be constructed for a given purpose.

The "pharmaco-genomics" is another interesting potential application of DNA microarrays in genomic suggested recently (Schena et al. 1996; Graves 1999). Because each individual has a slightly different genetic makeup, each will have a unique set of polymorphic sites; although these polymorphisms might not be sufficiently aberrant to cause disease, some of them would determine how each individual responds to a particular drug. One might know ahead of time that he or she would have an adverse reaction to a particular drug or that it would be ineffective for the disease by the proper DNA microarrays. Other genomic applications of DNA microarrays include the identification of criminals or of blood relatives, the tissue typing in organ-donor selection, and the studies on evolution and interspecies similarities (Richmond and Sommerville 2000). Biologists have rapidly recognized the importance of DNA microarray technology, as illustrated by ambitious genomic programs (Richmond et al. 1999; Wang *et al.* 2000) and by the establishment of core microarray facilities. Recently, the first large-scale analysis was performed to identify nitrate responsive genes using 5 524 unique cDNA clones representing approximately a quarter of the *Arabidopsis* genome (van Hal *et al.* 2000; Wang *et al.* 2000). Novel nitrate-induced genes were found and multiple responses to nitrate were observed at the transcript level. These demonstrated the power of such global investigation for gene discovery and for the analysis of regulatory networks. With the completion of genomic sequencing projects in several model organisms, genome-wide microarrays will become essential tools for discovering the function of genes.

Gene expression studies

Measuring transcript levels for thousands of genes in parallel is one of the more widespread applications of DNA chip technology. Microarrays for gene expression analysis were the first biological application of DNA chip technology. Both oligonucleotide and cDNA microarrays work well for transcript monitoring (Desprez *et al.* 1998; Durrant *et al.* 2000; Kehoe *et al.* 1999; Maleck *et al.* 2000; Matsumura *et al.* 1999; Roth *et al.* 1998; Rushton and Somssich 1998) One advantage of oligonucleotide microarrays for expression studies is that chips can be prepared directly from sequence databases, obviating the need for cumbersome clone handling and sample tracking. Another advantage of oligonucleotide microarrays is that transcripts from individual members of multi-gene families that share extensive sequence homology can be easily distinguished by synthesizing oligonucleotides to regions of non-identity. Microarrays of cDNAs possess some distinct advantages over oligonucleotide microarrays, including the ease of prototyping and data analysis, immediate accessibility to the research community, and the capacity to examine large numbers of novel cDNAs in gene discovery applications. Both oligonucleotide and cDNA microarrays will be widely used for gene expression analysis. Three strategies can be adopted when developing a DNA chip for gene expression studies. The first consists of strategically selecting genes that are known to play an important role in a particular biological pathway. The second strategy, which is restricted to cDNA microarrays, is to use clones from a library prior to sequence analysis (Rushton and Somssich 1998). The third strategy is to generate a chip with the complete expressed sequence content of an organism (Thiellement *et al.* 1999). Genome chips provide the best chances for discovering new interactions between metabolic and genetic pathways and for gaining functional insights into novel expressed sequences.

The recent advent of tools enabling the global analysis of gene expression coupled to the genome sequencing of model species like *Arabidopsis* or rice are dramatically changing the way experimentation is done and provide the research community with the hope of answering more general questions. One technological advance, DNA microarrays, is already becoming a standard tool for genome-wide monitoring of gene expression in animal studies and is starting to contribute to the field of plant biology. Models established by DNA microarrays serve to organize current information, relationship and hypothesis, and can be tre-

mendously helpful for testing new hypothesis, interpreting new observations, designing experiments, and predicting the lively effects of particular chemical, genetic or cellular perturbations. DNA microarrays are already being used to study how cells respond to environmental changes and stress through changing mRNA patterns. It has been suggested that this technology might be extended to studies of environmental toxicity caused by dioxin or mercury, by looking for subtle changes in gene expression 9, and to evaluate the many products resulting from combinatorial chemistry, which might not cause obvious changes in cellular appearance or behavior but could cause subtle metabolic changes that would show up when the mRNA was interrogated by an array. One day, it might be possible to control the life cycle of an animal or plant much more precisely or to find an efficient insecticide that does not affect other species based on DNA microarrays. Cancerous cells often have a number of unique characteristics, including loss of hetero-zygosity and fusion transcripts, and changes in tumor-suppressor genes (Cheung *et al.* 1999; DeRisi *et al.* 1997) or oncogenes. Some of these changes might be detected more easily through expression changes. It is likely that the activity of certain genes will be up-regulated and others down-regulated in actively dividing malignant cells. Arrays might eventually be used to predict whether a particular tumor would respond to a particular drug and to obtain an early indication of recurrence following remission or a seemingly effective therapy.

Biological signal and defense system

Recent progress in understanding biological signal and defense system has highlighted an interacting network of signaling pathways leading to the induction of numerous genes. The combination of expression data with other biochemical or metabolite measurements seems another promising approach. Induced defense has received a lot of attention and over the years a large number of genes encoding defense-related proteins have been identified. A vast majority of these genes are induced after the plants were attacked by diverse aggressors such as microbial pathogens, viruses, or insects. Understanding the signaling machinery that links the perception of the incoming enemy to specific changes in gene expression has been the focus of recent research and several key molecular components have been isolated in *Arabidopsis thaliana* (Glazebrook 1999). The emerging picture is that a complex network of interdependent signaling pathways convey the information on the nature of the aggressor and allows the plant to mount an appropriate defense response (Reymond *et al.* 1998). Plants have to deal with a vast range of pathogens and it is not known what proportion of the genome is allocated to defense.

The described transcript profiles during systemic acquired resistance (SAR), a defense reaction known to develop in systemic leaves after an initial pathogen attack of local leaves (Maleck *et al.* 2000) indicated that 4.3 % of the genes (300 out of 7 000) were involved in the SAR response. The metabolic profiling technique has also recently permitted the discovery and quantification of fatty acid-derived molecules that accumulate during wounding and pathogenesis (Vollenweider *et al.* 2000). The role of these molecules as biological regulators implicated in defense will be tested by microarray analysis, highlighting a potential application of transcript profiling methods for discovering the function of new metabolites. When the repertoires of transcripts and metabolites measured in a single experiment increase to genome-scale levels, the challenge will be to integrate these complex databases and to extract meaningful biological information. In order to fully understand complex defense responses, input from proteomic and metabolomic studies will be essential (Roessner *et al.* 2000; Trethwey *et al.* 1999).

Cell cycles regulation

A much larger fraction of cell cycle-modulated genes is in DNA synthesis, cell growth or cell division. Although there is a strong correlation between distinct experimental profiles and functional assignment, not all genes involved in DNA replication are expressed periodically in the cell cycle, and some gene that do not need to be cell cycle-regulated are transcribed in a periodic fashion (Tanaka *et al.* 2000; Velculescu *et al.* 1995). Studies of cell cycle regulation have focused on genes with cell cycle–specific functions. These genes whose functions are only needed for a part of the cycle are directly involved in DNA replication and mitosis. For some such genes, transcriptional regulation may be a matter of conserving resources. Genes needed for cycling are evidently not needed during the dormant period, but very much needed immediately afterward. Thus, cell cycle-regulated expression may ensure that necessary gene products are always available to cycling cells. In this regard, it is interesting to note that the purine-rich motif AAGAAAAA (Spellman *et al.* 1998) is thought to be important for response to glucose; this motif may be important in the switch from stationary phase to rapid growth, and we find similar motifs enriched in the promoters of several types of cell cycle-regulated genes. That is, these genes may be growth regulated as well as cell cycle regulated. Other genes with cell cycle-specific functions act as regulators or switches. It is not only important when exactly they are on but also when they are off. Transcriptional regulation of a gene controlling a switch can be central to its function (Velculescu *et al.* 1995; Graves 1999).

Cell cycle-regulated transcription can be used to build a structure in a highly controlled way. This can be illustrated with some parallels between the strategy of the cell for regulating DNA replication and its strategy for regulating differentiation. Transcriptional controls provide key components of the initiation complexes at certain times so that the complexes can be built in an orderly manner; however, the complexes cannot easily later be rebuilt at an inappro-

priate time, partly because the components are no longer available. Many cell cycle-regulated genes whose functions are essentially not cell cycle specific. The best single example of such a gene may be PMA1, encoding the major plasma membrane proton pump, a stable protein. The PMA1 function is essential, and although its function is required throughout the cell cycle (van Hal *et al.* 2000; Spellman *et al.* 1998), its transcription is strongly periodic. With the DNA microarrays analysis, Spellman *et al.* (1998) found 800 yeast genes whose transcripts oscillate through one peak per cell cycle. They defined these 800 genes by using an empirical model of cell cycle regulation, whose threshold was somewhat arbitrary. Below this threshold there may well be genes whose expression is truly periodic and whose periodicity might even have biological significance. They observed independently that their expression was affected by induction of Cln3p and Clb2p. Although the basis of the regulation of the remaining genes and some of the detailed behavior of some of the cyclin-dependent gene expression remains to be elucidated, there will be increasing value in genomic data sets as more of them accumulate and that together these will fully realize the promise of the genome sequencing projects (Graves 1999; Cheung *et al.* 1999).

Mechanisms of transcriptional regulation

When the information on the complete genome sequence is available, gene expression data can be used to identify new genomic sequence motifs that are over-represented in the genomic DNA in the vicinity of similarity behaving genes and set of co-regulated genes. The correlation between the presence of specific sequence motifs in promoter regions and gene expression patterns may be stronger than the correlation between functional categories and gene expression patterns. Spellman *et al.* (1998) had examined 800 genes for the binding sites of known cell cycle transcription factors by DNA microarrays. They found that 400 genes were good matches to known sites relevant to the phase of peak expression and 280 of these same genes showed a significant response to Cln3p or Clb2p induction. In addition, they identified as cell cycle regulated other sets of genes that form functional pathways and were known to be co-regulated or about whose regulation something is known. The DNA microarray data give us a partial picture of the logical circuitry of transcriptional controls in the cell cycle. A large number of genes are induced in G1 and S by the action of Cln3p-Cdc28p on MBF and SBF. Furthermore, by M they become repressed by the action of Clb2p-Cdc28p. At the same time, Clb2p-Cdc28p, acting through MCM1 1 SFF, induces its own constellation of genes 50. These genes include the important transcription factor Swi5p; once its transcription has been induced, it is allowed to enter the nucleus (van Hal *et al.* 2000; Rushton and Somssich 1998). The loss of Clb2-Cdc28 activity causes a collapse in the transcription of all Clb2p-dependent transcripts and allows Cln3p-Cdc28p to reactivate MBF and SBF to begin a new cell cycle. However, the oscillation in the genes expressed in M/G1 phase from an MCM1 site (the ECB) and what makes expression of these genes cell cycle regulated remain to be explained (Duggan *et al.* 1999; Rushton and Somssich 1998).

In addition, microarrays will help in the identification of genes whose expression is controlled by known transcription factors. Transcription factors could be over-expressed or silenced in transgenic plants and the effect on gene expression measured by microarray analysis. This was achieved successfully in yeast for defining target genes modulated by transcription factors involved in oxidative stress (Wang *et al.* 2000). Another particularly promising way of using microarrays for understanding the mechanisms of pathogenesis will be to compare the responses induced by various pests or microorganisms. For instance, an analysis of transcript profiles after challenging plants with different pathogens or stimuli might answer the question concerning host discrimination between pathogens and might help identifying transcript profiles among host responses. However, a larger effort will be necessary before pathogen specific-transcript profiles can be defined but this opens the perspective of being able to precisely diagnose plant diseases at the molecular level and will undoubtedly be of central importance for agriculture. As model genomes are sequenced, it will soon be possible to have the complete sets of genes of both the host and the pathogen on the same microarray, producing a unique molecular view of the interaction between the plant and its aggressor (DeRisi *et al.* 1997; Rushton and Somssich 1998).

Molecular bar coding and reverse genetics

Another experimental genomic activity in which DNA chips will play a central role is the characterization of populations of mutant organisms exposed to various selective pressures. The completion of the yeast genome sequencing project has catapulted the yeast genetics community into the post-genome era. One of the next logical steps for the yeast community is the systematic preparation of single gene deletion mutants corresponding to all 6 000 open reading frames (ORFs). Indeed, the ease with which yeast genes can be targeted by homologous recombination is a central advantage of this model system. Shoemaker (Simpson *et al.* 2000) has proposed a new strategy for screening large populations of knock-out mutants in parallel. Their strategy consists of introducing unique molecular sequences or 'bar codes' into each of the 6 000 ORFs in the yeast genome. These unique 20-mers can then be used for parallel hybridization-analysis with oligonucleotide microarrays. In this strategy, a pool of yeast strains containing individual bar codes for all 6 000 genes is subjected to a selective pressure. Samples of cells growing under selective conditions are taken at incremental times during the course of the experiment and the bar-code sequences are labeled by multiplex PCR with fluorescent primers.

Each pool of fluorescent amplicons is then hybridized to an oligonucleotide microarray containing sequences complementary to each of the amplified bar codes (Reymond *et al.* 1998; Trethewey *et al.* 1999).

Comparative analysis of fluorescent intensities at each bar code position over time provides a quantitative measure of the fitness of each strain under a given selective pressure. Correlation between strain disappearance and selective pressure allow global functional analysis of yeast gene function. Because genetic bar coding is not yet feasible in higher organisms, systematic searches for mutant alleles will be required for genetic analysis in these systems. In the case of plants, Arabidopsis is likely the organism of choice for these efforts. Feldmann and collaborators have proposed the use of expressed sequence tags (ESTs) as a global means of identifying "insertion elements" in genes of interest (van Hal *et al.* 2000; Aharoni *et al.* 2000). In this approach, PCR is used to screen pools of Arabidopsis lines bearing insertion elements at random locations. Lines bearing a mutational insert in a gene of interest yield a specific amplicon in the PCR amplification. Although this method can be performed for all of sequenced genes of Arabidopsis, it is both labor intensive and costly in terms of primer synthesis. Studies have shown that interlaced asymmetric PCR can be used to generate products of plant DNA / T-DNA insert junctions (van Hal *et al.* 2000; Lockhart *et al.* 1996). Hybridization of PCR amplicons to microarrays of expressed sequences could be used to speed the identification of mutant lines of Arabidopsis.

DNA microarrays and proteomics

The elucidation of protein/protein interactions within cells, as well as the identification of proteins that bind small ligands, is another area in which DNA chips could significantly increase the rate of discovery. Bartel (van Hal *et al.* 2000; Wodicka *et al.* 1997) have demonstrated that a 'protein-linkage' map can be created using genomic sequence information. The authors correctly suggest that the yeast two-hybrid system is probably the best tool available for the systematic determination of protein-protein interactions in complex organisms. The two-hybrid system uses two fusion proteins to activate the transcription of reporter genes in yeast (Eisen *et al.* 1998; Blackstock *et al.* 1999). The first fusion protein contains a DNA binding domain fused to a protein of interest, while the second is an acidic transcriptional activation domain fused to a second protein of interest. Specific interactions between the two chimeric proteins leads to transcriptional activation the reporter genes, which is easily scored with either color-based assays or by auxotrophic complementation. In the conventional two-hybrid approach, the identity of interacting proteins is confirmed by sequence analysis of each clone identified in the yeast assay.

As an alternative to conventional DNA sequencing, it is possible to use chip hybridization to identify the genes involved in protein-protein interactions. In the case where entire genome sequences are available (e.g., yeast), DNA chips can be used for massive, parallel gene resequencing. With the chips, it would be possible to rapidly identify all of the clones whose encoded sequences interact in the two-hybrid assay. In this experimental design, PCR would be used to amplify and label each cDNA insert that encodes an interacting protein. Hybridization to genome chips would allows identification of the all of the genes involved in protein-protein interaction in a single hybridization. Phage presentation libraries are also amenable to DNA chip-based detection systems. Phage presentation utilizes fusion proteins encoded by chimeric sequences of bacteriophage viral coat proteins and genes of interest (Blackstock & Weir 1999). As in the case of the two-hybrid system, cDNA libraries encoding fusion proteins are created. Phage display libraries can be made for plant species by fusing coding sequences to plant virus coat proteins (van Hal *et al.* 2000). Individual clones in the library can be selected by 'panning' with any ligand of interest. If this ligand is immobilized to an inert support, the bacteriophage clones that bind can be purified by elution. The bacteriophage particles can then be used as templates for PCR, with a set of fluorescent-labeled primers that flank the cDNA inserts. The cDNA fragments in this enriched population could then be characterized by microarray hybridization.

Functionality of food component

The DNA microarray technologies can be used in the safety assessment of genetic modification of food plants. This should not only be based on the evaluation of the newly introduced trait, but also on possible unintended side effects resulting from the genetic alteration. Presently, the latter is done by comparison of a limited number of individual macro- and micronutrients as well as anti-nutritional factors, including natural toxins, between the transformed line and its traditional counterpart. However, this approach has its limitations, since it is unknown whether the currently investigated components are the only and most important factors to screen with respect to food safety. This type of analysis therefore provides only limited information on the potential effects on human health. DNA microarray technology provides the opportunity to screen for unintended changes in expression of large numbers of genes in an unbiased manner. Van Hal *et al.* (2000) and Tavazoie *et al.* (1999) had tested this approach for tomato. They have constructed tomato cDNA libraries enriched for cDNAs preferentially expressed in either green or red tomato. Arrays containing these sequences as well as target sequences of known genes will be used for the systematic comparison of control and genetically modified tomatoes. This will provide information on the natural variation in gene expression of tomato at certain stages of ripening as well as specific changes due to genetic alteration (Tavazoie *et al.* 1999)

One way to identify relevant biological functions is by large-scale expression analysis, since a large number of characters can be tested in parallel and a large number of different exposures can be compared. We are focusing on effects of food components on the human intestine (Bassett *et al.* 1999). Differentially expressed genes will be characterized by sequencing and further expression analysis in a number of different intestinal cell lines as well as in vivo. DNA-microarray-assisted gene expression analysis offers a powerful tool to identify the genes that are affected in a certain mutant. Those genes will be related to the nature of the mutation and may give an important clue in elucidating the function of the mutated gene in transgenic animals or plants (Bassett *et al.* 1999).

Perspective

One of the challenges facing the research community will be to deal with the flow of data generated by whole genome expression studies (Bassett *et al.* 1999; Ermolaeva *et al.* 1998). The database developed at Stanford University host data from multiple microarray analyses in different organisms and offers several useful query tools. It will be important in the future to be able to compare expression data across many different experimental, geographical or technical platforms. In the next few years, DNA microarrays will certainly become a standard tool in each laboratory. We can foresee two ways of using the technology with the aim of answering different types of questions (Fig. 3). First, microarrays containing a representation of the whole plant genome will be served to identify the expression pattern of genes of unknown function, to define specific sets of genes responding to various stresses or stimuli, to provide a global view on metabolic processes, and to assist in comparing wild type and mutant organisms. Because the production and routine use of whole genome-microarrays might be financially too demanding for most research groups, genomic centers providing access to large microarrays might develop further and allow the screening for genes of specific interest. Both DNA microarray and Affymetrix chip technologies are complementary and suited for the production and analysis of whole genome based arrays (Reymond *et al.* 2000).

These approaches will be quite useful for a deep and thorough analysis of the expression patterns of hundreds of genes and would be affordable for most research groups (Tavazoie *et al.* 1999; Ermolaeva *et al.* 1998). Such studies may include more detailed characterization of expression patterns, including replication of multiple experiments and time-course analyses. The strategy of fabricating custom arrays tailored to a specific biological question has the advantage of being easier to control at the production side while reducing the amount of data to process and integrate. Another source of candidate genes involved in specific plant responses and which might also constitute boutique microarrays will come from differential screening methods, such as differential display and RNA finger-printing (cDNA-AFLP), or might simply be constituted on the basis of literature search or in silico-analyses of SAGE and ESTs databases. Whole genome-microarray analyses will be serviced by core facilities and will provide users with the possibility of performing global analyses of gene expression to study regulatory processes or to discover the function of unknown genes. Large-scale microarrays will also help in defining sets of genes that will be chosen to fabricate less expensive custom microarrays within each laboratory. These microarrays will be tailored to more specific research projects and will be more useful for routine-based transcript profiling (Fig. 3). Dedicated microarrays containing a well-defined set of defense-related genes have already demonstrated their utility for the study of wound- and insect inducible gene expression and the involvement of signal molecules in the wound and pathogen responses (Reymond *et al.* 2000).

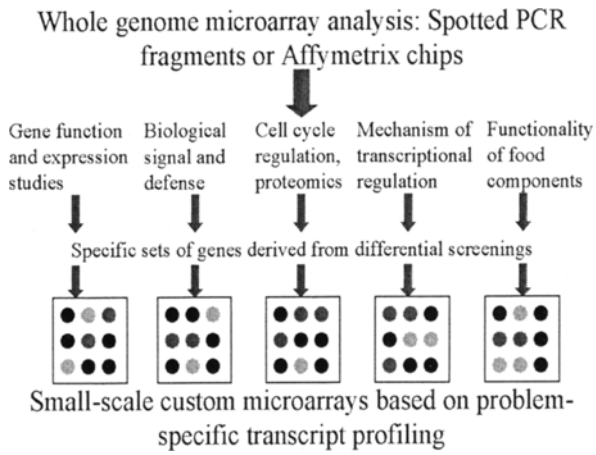

Fig. 3 Schematic overview on the potential use of DNA microarrays in the future.

Large-scale microarrays will help in defining sets of genes that will be chosen to fabricate less expensive custom microarrays within each laboratory. These microarrays will be more useful for routine-based transcript profiling.

Conclusions

DNA microarrays will substantially increase the speed at which differential gene expression can be analyzed and gene functions can be elucidated. Genome sequencing programs have already produced large amounts of sequence data. Components or pathway engineering is expected to accelerate research and improve knowledge in the fields of cellular and molecular biology. To fully fulfill these expectations, further improvement of the technology with respect to reproducibility, speed, cost and sensitivity will be needed. Consequently, sufficient attention should be paid to the development of biological model systems, which will facilitate further optimization that is asked for in its various applications. Information obtained from genom-

ics, large-scale expression analysis, proteomics and metabolite profiling will be invaluable to identify gene functions, pathways and interactive cellular physiology. DNA microarrays containing a full animal or plant genome might soon be available in more model systems and will certainly contribute to a precise knowledge on all events occurring during growth, development, differentiation, and pathogenesis and will be crucial for the discovery of gene function. Small-scale custom arrays with dedicated sets of genes might also prove to be useful for a deep and thorough analysis of the biological processes that take place in a cell. It is expected that DNA microarrays will greatly help in studying these complex interactions in cells or organisms.

Acknowledgments

We are very grateful to Professor R. Martienssen, Professor C. Koncz, and Professor R. Sederoff for encouragement and support.

References

Aharoni, A., Keizer, L.C., Bouwmeester, H.J. *et al.* 2000. Identification of the SAAT gene involved in strawberry flavor biogenesis by use of DNA microarrays [J]. Plant Cell, **12**: 647-662.

Alwine, J.C., Kemp, D.J., Stark, G.R. 1997. Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes [J]. Proc. Natl. Acad. Sci. USA, **74**: 5350–5354.

Baldwin, D., Crane, V., Rice, D. 1999. A comparison of gel-based, nylon filter and microarray techniques to detect differential RNA expression in plants [J]. Curr. Opin. Plant Biol., **2**: 96-103.

Bassett, D.E. Jr, Eisen, M.B., Boguski, M.S. 1999. Gene expression informatics – it's all in your mine [J]. Nat. Genet., **21**: 51–55.

Blackstock, W.P., Weir, M.P. 1999. Proteomics: quantitative and physical mapping of cellular proteins [J]. Trends Biotechnol., **17**: 121-127.

Chee, M., Yang, R., Hubbell, E., *et al.* 1996. Accessing genetic information with high-density DNA arrays [J]. Science, **274**: 610-614.

Cheung V.G., Morley M., Aguilar F. *et al.* 1999. Making and reading microarrays [J]. Nat. Genet., **21**: 15-19.

Chu, S., DeRisi J., Eisen M. *et al.* 1998. The transcriptional program of sporulation in budding yeast [J]. Science, **282**: 699-705.

DeRisi, J.L., Iyer, V.R., Brown P.O. 1997. Exploring the metabolic and genetic control of gene expression on a genomic scale [J]. Science, **278**: 680-686.

Desprez, T., Amselem, J., Caboche, M *et al.* 1998. Differential gene expression in Arabidopsis monitored using cDNA arrays [J]. Plant J., **14**: 643-652.

Duggan, D.J., Bittner, M., Chen, Y. *et al.* 1999. Expression profiling using cDNA microarrays [J]. Nat. Genet., **21**. S10-S14

Durrant, W.E., Rowland, O., Piedras, P. *et al.* 2000. cDNA-AFLP reveals a striking overlap in race-specific resistance and wound response gene expression profiles [J]. Plant Cell, **12**: 963-977.

Edman, C.F., Raymond, D.E., Wu, D.J. *et al.* 1997. Electric field directed nucleic acid hybridization on microchips [J]. Nucl. Acids Res., **25**: 4907-4914.

Eisen, M.B., Spellman, P.T., Brown, P.O. *et al.* 1998. Cluster analysis and display of genome-wide expression patterns [J]. Proc. Natl. Acad. Sci. USA, **95**: 14863-14868.

Ermolaeva, O., Rastogi, M., Pruitt, K.D. *et al.* 1998. Data management and analysis for gene expression arrays [J]. Nat. Genet., **20**: 19-23.

Fodor, S.P.A., Read, J.L., Pirrung, M.C. *et al.* 1991. Light-directed, spatially addressable parallel chemical synthesis [J]. Science, **251**: 767-773.

Geiger, A., Lester, A., Kleiber, J. *et al.* 1998. PNA array technology in molecular diagnostics [J]. Nucleotides, **17**: 1717-1724.

Glazebrook J. 1999. Genes controlling expression of defense responses in Arabidopsis [J]. Curr. Opin. Plant Biol., **2**: 280-286.

Graves, D.J. 1999. Powerful tools for genetic analysis come of age [J]. Trends Biotechnol., , **17**: 127-134.

Iyer, V.R., Eisen, M.B., Ross, D.T. *et al.* 1999. The transcriptional program in the response of human fibroblasts to serum [J]. Science, **283**: 83- 87.

Jordan, B.R. 1998. Large-scale expression measurement by hybridization methods: from high-density membranes to 'DNA chips' [J]. J. Biochem., **124**: 251-258.

Kehoe, D.M., Villand, P., Somerville, S. 1999. DNA microarrays for studies of higher plants and other photosynthetic organisms [J]. Trends Plant Sci., **4**: 38-41.

Lemieux, B., Aharoni, A., Schena, M. 1998. Overview of DNA chip technology [J]. Mol. Breed., **4**: 277-289.

Lennon, G.G., Lehrach, H. 1991. Hybridization analyses of arrayed cDNA libraries [J]. Trends Genet., **7**: 314-317.

Liang, P., Pardee, A.B. 1992. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction [J]. Science, **257**: 967-971.

Lipshutz, R.J., Fodor, S.P.A., Gingeras, T.R., *et al.* 1999. High density synthetic oligonucleotide arrays [J]. Nat. Genet., **21**: S20–S24.

Lockhart, D.J., Dong, H., Byrne, M.C. *et al.* 1996. Expression monitoring by hybridization to high-density oligonucleotide arrays [J]. Nat. Biotechnol., **14**: 1675-1680.

Maleck, K., Levine, A., Eulgem, T. *et al.* 2000. The transcriptome of Arabidopsis during systemic acquired resistance [J]. Nat. Genet.,**26**: 403-410.

Matsumura, H., Nirasawa, S., Terauchi, R. 1999. Transcript profiling in rice (*Oryza sativa* L.) seedlings using serial analysis of gene expression (SAGE) [J]. Plant J., **20**: 719-726.

McGall, G., Labadie, J., Brock, P. *et al.* 1996. Light-directed synthesis of high-density oligonucleotide arrays using semiconductor photoresists [J]. Proc. Natl. Acad. Sci. USA, **93**: 13555-13560.

Proudnikov, D., Timofeev, E., Mirzabekov, A. 1998. Immobilization of DNA in polyacrylamide gel for the manufacture of DNA and DNA-oligonucleotide microchips [J]. Anal. Biochem., **259**: 34-41.

Reymond, P., Farmer, E.E. 1998. Jasmonate and salicylate as global signals for defense gene expression [J]. Cur. Opin. Plant Biol., **1**: 404-411.

Reymond, P., Weber, H., Damond, M. *et al.* 2000. Differential gene expression in response to mechanical wounding and insect feeding in Arabidopsis [J]. Plant Cell, **12**: 707-720.

Richmond, C.S., Glasner, J.D., Mau, R. *et al.* 1999. Genome-wide expression profiling in *Escherichia coli* K-12 [J]. Nucleic Acids Res., **27**: 3821-3835.

Richmond, T., Somerville, S. 2000. Chasing the dream: plant EST microarrays [J]. Curr. Opin. Plant Biol., **3**: 108-116.

Roessner, U., Wagner C., Kopka J., *et al.* 2000. Simultaneous analysis of metabolites in potato tuber by gas chromatography-mass spectrometry [J]. Plant J., **23**: 131-142.

Rogers, Y.H., Jiang-Baucom, P., Huang, Z.J. *et al.* 1999. Immobilization of oligonucleotides onto a glass support via disulfide bonds: a method for preparation of DNA microarrays [J]. Anal. Biochem., **266**: 23-30.

Roth, F.P., Hughes, J.D., Estep, P.W. *et al.* 1998. Finding DNA regulatory motifs within unaligned noncoding sequences clustered by whole-genome mRNA quantitation [J]. Nat. Biotechnol., **16**: 939-945.

Rushton, P.J., Somssich, I.E. 1998. Transcriptional control of plant genes responsive to pathogens [J]. Curr. Opin. Plant Biol., **1**: 311-315.

Schena, M., Shalon, D., Davis, R.W. *et al.* 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray [J]. Science, **270**: 467-470.

Schena, M., Shalon, D., Davis, R.W., *et al.* 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray [J]. Science, **270**: 467-470.

Schena, M., Shalon, D., Heller, R. *et al.* 1996. Parallel human genome analysis: microarray-based expression monitoring of 1000 genes [J]. Proc. Natl. Acad. Sci. USA, **93**: 10614-10619.

Schena, M., Shalon, D., Heller, R., *et al.* 1996. Parallel human genome analysis: microarray- based expression monitoring of 1000 genes [J]. Proc. Natl. Acad. Sci. USA, **93**: 10614-10619.

Simpson, A.J.G., Reinach, F.C., Arruda, P. *et al.* 2000. The genome sequence of the plant pathogen *Xylella fastidiosa* [J]. Nature, **406**: 151-159.

Sosnowski, R.G., Tu, E., Butler, W.F. *et al.* 1997. Rapid determination of single base mismatch mutations in DNA hybrids by direct electric field control [J]. Proc. Natl. Acad. Sci. USA, **94**: 1119-1123.

Spellman, P.T., Sherlock, G., Zhang, M.Q. *et al.* 1998. Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization [J]. Mol. Biol. Cell., **9**: 3273-3297.

Tanaka, T.S., Jaradat S.A., Lim M.K., *et al.* 2000. Genome-wide expression profiling of mid-gestation placenta and embryo using a 15,000 mouse developmental cDNA microarray [J]. Proc. Natl. Acad. Sci. USA, **97**: 9127-9132.

Tavazoie, S., Hughes, J.D., Campbell, M.J. *et al.* 1999. Systematic determination of genetic network architecture [J]. Nat. Genet., **22**: 281–285.

The *C. elegans* Sequencing Consortium. 1998. Genome sequence of the nematode *C. elegans*: a platform for investigating biology, Science, **282**: 2012-2018.

Thiellement, H., Bahrman, N., Damerval, C. *et al.* 1999. Proteomics for genetic and physiological studies in plants [J]. Electrophoresis, **20**: 2013-2026.

Trethewey, R.N., Krotzky A.J., Willmitzer L. 1999. Metabolic profiling: a Rosetta Stone for genomics? [J]. Curr. Opin. Plant Biol., **2**: 83–85.

van Hal N.L., Vorst O., van Houwelingen A.M. *et al.* 2000. The application of DNA microarrays in gene expression analysis [J]. J. Biotechnol., **78**: 271-280.

Velculescu, V.E., Zhang, L. 1995. Vogelstein, B., et al., Serial analysis of gene expression [J]. Science, **270**: 484-487.

Vollenweider, S., Weber, H., Stolz, S., Farmer, E.E. 2000. Fatty acid ketodienes and fatty acid ketotrienes: Michael addition acceptors that accumulate in wounded and diseased Arabidopsis leaves [J]. Plant J., **24**: 467-476.

Wang, R., Guegler K., LaBrie S.T. *et al.* 2000. Genomic analysis of a nutrient response in Arabidopsis reveals diverse expression patterns and novel metabolic and potential regulatory genes induced by nitrate [J]. Plant Cell, **12**: 1491-1510.

Watson, A., Mazumder, A., Stewart, M., *et al.* 1998. Technology for microarray analysis of gene expression [J]. Curr. Opin. Biotechnol., **9**: 609-614.

White, K.P., Rifkin S A., Hurban P. *et al.* 1999. Microarray analysis of *Drosophila* development during metamorphosis [J]. Science, **286**: 2179-2184.

Wittung, P., Nielsen, P., Norden, B. 1996. Direct observationof strand invasion by peptide nucleic acid (PNA) into double-stranded DNA [J]. J. Am. Chem. Soc., **118**: 7049-7054.

Wodicka, L., Dong, H., Mittman, M. *et al.* 1997. Genome-wide expression monitoring in *Saccharomyces cerevisiae* [J]. Nat. Biotechnol., **15**: 1359-1367.